David Happel ● CSE3000 ● 25-06-2020 ● Stavros Makrodimitris ● Arman Naseri Jshfari ● Tom Vierling ● David Tax ● Marco Loog
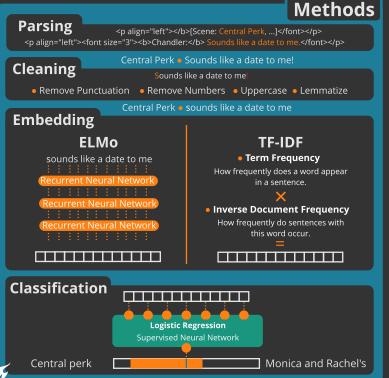
● Extracting location context from transcripts ●

Research

Using transcripts from the series F·R·I·E·N·D·S
To what extent can we predict at which location a line was said using feature extraction techniques?

Methods

Parsing

`<p align="left"></b>[Scene: Central Perk, ...]</font></p>`
`<p align="left"><font size="3"><b>Chandler:</b> Sounds like a date to me.</font></p>`

Central Perk ● Sounds like a date to me!

Cleaning

Sounds like a date to me!

● Remove Punctuation ● Remove Numbers ● Uppercase ● Lemmatize

Central Perk ● sounds like a date to me

Embedding

ELMo	TF-IDF
sounds like a date to me	**Term Frequency**
Recurrent Neural Network	How frequently does a word appear in a sentence.
Recurrent Neural Network	✕
Recurrent Neural Network	**Inverse Document Frequency**
	How frequently do sentences with this word occur.
	=

Classification

Logistic Regression
Supervised Neural Network

Central perk ─────── Monica and Rachel's

Results

Accuracy
Using 5-fold cross-validation

ELMo	TF-IDF
58.8%	**66.2%**

Effective Cleaning

	ELMo	TF-IDF
● Lemmatiztion	✕	✕
● Rem. Numbers	✓	✕
● Rem. Punctuation	✓	✓

Minimum Word Filter

● Removing short sentences improves performance.

TF-IDF word associations

Words that are strongly associated with either location by the classifier.

Central perk   upstairs ● dollar ● spontaneous ● cops ● Russ

Monica and Rachel's   beer ● neighborhood ● kitchen ● breast ● tasted

D.V.Happel@student.tudelft.nl

TUDelft